Identification of Stipules reduced, a leaf morphology gene in pea (Pisum sativum)

Moreau, Carol; Hofer, Julie Margareth; Eléouët, Morgane ; Sinjushin, Andrey; Ambrose, Mike ; Skøt, Kirsten; Blackmore, Tina; Swain, Martin; Hegarty, Matthew; Balanza, Vincent; Ferrándiz, Cristina; Ellis, T. H. Noel

Published in:
New Phytologist

DOI:
10.1111/nph.15286

Publication date:
2018

Citation for published version (APA):
Moreau, C., Hofer, J. M., Eléouët, M., Sinjushin, A., Ambrose, M., Skøt, K., Blackmore, T., Swain, M., Hegarty, M., Balanza, V., Ferrándiz, C., & Ellis, T. H. N. (2018). Identification of Stipules reduced, a leaf morphology gene in pea (Pisum sativum). *New Phytologist*, *220*(1), 288-299. https://doi.org/10.1111/nph.15286

Identification of *Stipules reduced,* a leaf morphology gene in pea (*Pisum sativum*).

Carol Moreau[2], Julie M.I Hofer [1,5], Morgane Eléouët[1], Andrey Sinjushin[3], Mike Ambrose[2], Kirsten Skøt[1], Tina Blackmore[1], Martin Swain[1], Matthew Hegarty[1], Vicente Balanzà[6], Cristina Ferrándiz[6], and T.H. Noel Ellis[1,4]

[1] IBERS, Aberystwyth University, Plas Gogerddan, Aberystwyth, SY23 3EE, UK
[2] John Innes Centre, Norwich, NR4 7UH, UK
[3] Genetics Department M.V. Lomonosov Moscow State University, 119991, Moscow, Russia
[4] School of Biological Sciences, University of Auckland, 1142, New Zealand
[5] Faculty of Education and Social Work, University of Auckland, 1023, New Zealand
[6] Instituto de Biología Molecular y Celular de Plantas, CSIC, 46022, Valencia, Spain

Corresponding Author	Noel Ellis
e-mail (preferred)	thnoelellis@gmail.com
alternative e-mail	tell822@aucklanduni.ac.nz
Telephone	+64 27 636 7531
Word count Total	6,467
	Word count by section: Introduction (498) Materials and Methods (1,101), Results (2,644), Discussion (2,187), and Acknowledgements (37)
Figures	7 Figures, 3 colour
Tables	1
Supporting information	Four files in total. File 1 has 10 supplementary figures Fig. S1 - Fig. S10. File 2 is Methods S1 containing a Python script and Files 3 and 4 are data files, Notes S1 and Notes S2, with sequence related information.

Summary

- Pea (*Pisum sativum* L.) is one of relatively few genetically amenable plant species with compound leaves. Pea leaves have a variety of specialised organs: leaflets, tendrils, pulvini and stipules which enable the identification of mutations that transform or affect distinct parts of the leaf. Characterisation of these mutations offers insights into the development and evolution of novel leaf traits. The previously characterised morphological gene *Cochleata*, conferring stipule identity, was known to interact with *Stipules reduced*, which conditions stipule size in pea, but the *Stipules reduced* gene remained unknown.
- Here we analysed Fast Neutron irradiated pea mutants by Restriction site Associated DNA sequencing.
- We identified *Stipules reduced* as a gene encoding a C2H2 zinc finger transcription factor that is regulated by *Cochleata*. *Stipules reduced* regulates both cell division and cell expansion in the stipule.
- Our approach shows how systematic genome-wide screens can be used successfully for the analysis of traits in species for which whole genome sequence is not available.

Key words

Pea (*Pisum sativum*) - leaf - stipule – RAD sequencing - Fast Neutron – mutant - C2H2 zinc finger

Introduction

Leaves have a variety of functions, notably photosynthesis, transpiration and canopy shading. The shape of the leaf impacts the energy efficiency and water economy of the plant, thus the genetics underlying variation in leaf form is of interest. In pea, a suite of leaf development mutants affecting pattern formation is known (Marx, 1987); these mutations affect the arrangement or identity of organs on the compound pea leaf. Mutations in six of these the genes have been characterised at the molecular level (Chen *et al*., 2012, Couzigou *et al*., 2012, Sainsbury *et al*., 2006, Zhuang *et al*., 2012 Hofer *et al*., 2009, 1997) and for each, at least one null allele has been generated by Fast Neutron (FN) bombardment in the line JI2822 (Domoney *et al*., 2013). These mutants offer a convenient opportunity for studying gene interactions in a single genetic background.

Most of the 28 mutant alleles characterised so far in this population are deletions of the entire gene; the deletion end-points are uncharacterised because they are far from the known sequence. Exceptions are two small deletions, one (of *a2*) is 22bp (Hellens *et al*., 2010) and the other (of *apu*) is 1.4kb (Chen *et al*., 2012) and one *b* allele is known to be a structural rearrangement (Moreau *et al*., 2012). These data suggest that the FN deletions are often large with respect to the size of a gene and are therefore amenable to systematic searches for deleted sequences. Deletion mutant alleles in this genetic background have been identified using AFLP markers (Hofer *et al.*, 2009); here we investigated whether AFLP markers could be replaced by restriction-site-associated (RAD) sequence markers (Miller *et al*., 2007, Baird *et al*., 2008).

We focussed attention on *Stipules reduced* (*St*; Pellew & Sverdrup 1923), a classical leaf morphology gene not yet cloned. The unusually large stipules of pea are replaced by smaller organs in the *st* mutant, more typical of close legume relatives in the genera *Lathyrus*, *Vicia* and *Lens*. The *st* stipules are also simpler in form than the wild type, lacking the serrated basal frill or mantle. In contrast to *Coch* (Couzigou *et al.*, 2012), the *St* gene is not required for stipule identity, rather it is a determinant of organ size, due to the early loss of marginal meristem activity (Meicenheimer et al., 1983). Large stipules are rare in legumes, but are found in *Pisum* and a few closely related *Lathyrus* species, notably *L.aphaca*, and in several other more distantly related taxa (Lewis *et al*. 2005). Whether stipule size has any adaptive significance is unknown, but the large stipules of pea enclose the developing shoot apical meristem and may protect it from frost damage or insect herbivory, or create a microclimate buffering the apex against changes in humidity or temperature. If there is some adaptive significance this could be reflected in sequence divergence parameters associated

with genes regulating stipule size. The isolation and characterisation of *St* as a C2H2 zinc finger gene will enable such studies in future.

Materials and Methods

Plant material

The line JI2822, a Recombinant Inbred Line (RIL) derived from a cross between JI15 and JI399, which has been described elsewhere (Lewis *et al*., 2005, Hofer *et al*., 2009), was mutagenised using Fast Neutrons from the ^{252}Cf facility at Oak Ridge National Laboratory, USA (Domoney *et al.*, 2013). Among the M2 progeny from these lines one mutant (FN2122/2) had a phenotype resembling *st* and a complementation test showed that it was allelic to *st.* Two additional mutant lines (FN1889/3 and FN2002/7) resembling, and allelic to, *creep* (Sidorova, 1975) were used as controls. Individual plants here designated FN1889-BC3, FN2002-BC3 and FN2122-BC4 were obtained by back crossing to JI2822 selections from M2 families that exhibited either the *creep* or *st* mutant phenotypes, followed by selfing and repeating this for three or four backcross cycles (as indicated by BC3 or BC4). Plants were finally selfed to extract homozygous mutants. M2 lineages have been estimated to carry an average of seven independent deletions (Domoney *et al*., 2013). After three cycles of backcrossing this is expected to be reduced to a single deletion.

The cross between the cultivars Flagman and Filby was generated at Lomonosov Moscow State University. The *st* mutant lines; JI17, JI132, JI143, JI924, JI1201, JI2160, JI2653, JI3528, JI3529, JI3530 JI3531 and JI3537 were obtained from the John Innes Pisum Germplasm collection, as were the wild type lines JI813 (cv Vinco), JI2822, JI3132 (cv Auralia), JI3538 (cv Paloma) and JI3539 (cv Virtus). The mutants FN2122 (*st*) and FN3185 (*coch*) are available from the John Innes Pisum Germplasm collection as JI3604 and JI3596 respectively.

Nucleic acid preparation

DNA was prepared as described by Ellis *et al.*, (1984), and RNA was prepared from seedling apices as described by Hofer *et al*., (2009).

Genetic Mapping

Genetic markers were analysed in the JI281xJI399 recombinant inbred population by simple matching with respect to existing markers. The identity of individual DNA samples prepared from this population was verified with PDR1 SSAP markers with the +TT primer combination (Ellis *et al.*, 1998), using fluorescent primers essentially as described by Knox *et al*., (2009).

RAD tags and sequencing

PstI digested DNA from four single plants (an FN2122-BC4 individual representing the mutant *st* allele; individuals of the two *creep* lines and JI2822, all representing the wild type *St* allele) was ligated to a RAD adaptor, sheared to the size range 100 - 600 nt and then ligated to the second adaptor prior to selective amplification as previously described (Domoney *et al.*, 2013). We used the method of Baird *et al*., (2008) with PstI instead of SbfI (or EcoRI) digestion; the PstI recognition site is internal to that of SbfI and the two enzymes leave the same overhang, so standard RAD sequencing primers were used for the sequencing reactions (Miller *et al*., 2007). Sequence reads were compiled into unique tags using FASTX Collapser (downloaded from http://hannonlab.cshl.edu/fastx_toolkit/) with the read depth per tag noted in the tag identifier. JI2822 was used as a reference and a Python script (Methods S1) was used to collate identical tags. This gave lists of tags of known read depth from each individual. Tags present in JI2822 but absent from the mutants were identified using Excel.

DNA sequence analysis

DNA sequence was obtained directly from an Illumina HiSeq 2500. Sequence characterisation of specific loci in pea and *Lathyrus* samples was performed using a BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems) using the primers listed in Notes S1 (for *Rca*) and S2 (for *St*). Pea and Lathyrus DNA sequences are deposited in Genbank with the accession numbers MF033127-MF033135.

Sequences related to *St* were obtained by BLAST analysis in the Legume Information System database (Dash *et al*., 2016, https://legumeinfo.org) and the pea gene atlas (Alves-Carvalho *et al*., 2015, http://bios.dijon.inra.fr/FATAL/cgi/pscam.cgi). For phylogenetic analysis those predicted amino acid sequences with an E value $< 10^{-20}$ over at least 40% of the sequence were selected for alignment using MUSCLE and the phylogenetic tree was constructed using PROML of the PHYLIP package with 1000 bootstraps. *RBE* was obtained from The Arabidopsis Information Resource (https://www.arabidopsis.org/index.jsp). The tree was drawn using DarWin5 (Perrier *et al*., 2003).

Quantitative PCR

cDNA was synthesized from total RNA prepared as described by Hofer *et al.*, (2009) using primers described in Notes S2. A Roche LightCycler 480 was used for qPCR experiments and amplicons were detected with SYBR green. Three biological and three technical replicates were used for each genotype. Reaction volumes were 15 µl with 0.25 µM primer. The cycling regime was 95°C for 5min, followed by 45 cycles of 95°C for 20s, 62°C for 15s, and 72°C for 15s. Analysis was done using the ΔΔCt method (Livak & Schmittgen,

2001) and corrected for primer efficiencies (Pfaffl, 2004). The average Cp value of the nine reactions was normalized relative to Actin.

Phloroglucinol staining

The nodal vasculature of *St* vs *st* plants was investigated using phloroglucinol staining. F2 progeny from a cross between cultivars Flagman (*St*) and Filby (*st*) were sown in a glasshouse. Nodes (4th and above) with stipules and short portions of petiole were excised from individual plants and fixed in 70% ethanol. After storing in ethanol for 2 months, nodes were soaked in 0.1 M phloroglucinol in 96% (w/v) ethanol for 12 h and then transferred to concentrated HCl for 1 h. After washing specimens in distilled water, the acid-macerated epidermis was partially removed. Images were captured from an Olympus SZ61 stereomicroscope (Olympus Corporation, Japan) using an ES-Experts 6600 digital camera (ES-Experts, Russia).

Epidermal cell area estimations

Epidermal cell replicas were made from the abaxial (lower) surface of *St* and *st* segregants of the Flagman x Filby cross using nail polish. These were examined microscopically (Nikon Eclipse Ci - Nikon, Japan) and photographed with a digital camera (Nikon DS-Vi1). Measurements were carried out on digital images using ImageJ 1.45m software (National Institute of Health, USA). Four zones were selected on the stipule lamina with respect to position on the proximodistal axis (proximal or distal) and the mediolateral axis of the leaf as a whole, where medial is closer to the petiole and lateral is away from the petiole. Thirty cells were measured in each zone. Statistical analysis was carried out using Statistica 8 (Statsoft, USA).

***In situ* hybridisation**

An RNA *in situ* hybridisation probe was derived from an *St* PCR product using primers p5 & p6 (Notes S2) using a T7 promoter in the reverse primer. *St* RNA was labelled with digoxigenin. Wax embedded sections were prepared from apices of JI2822 and the isogenic lines FN3185 (*coch* mutant) and FN2122 (s*t* mutant). *In situ* hybridisations were carried out according to Balanzà *et al*., (2018).

Results

Identification of a candidate gene.

Four genotypes were analysed; JI2822 (wild type *St*), FN2122/2 (*st*) as illustrated in Fig. 1 and Fig. S1, and two *creep* mutants FN1889/3 and FN2002/7 (both wild type *St*), the latter

used as independent controls. The *creep* mutant phenotype is not relevant and will not be discussed further.

A preliminary AFLP screen, following the method of Hofer *et al*., (2009), comparing JI2822 (wild type *St*) with FN2122/2 (*st*) identified a PCR product presence/absence polymorphism that distinguished these two genotypes (Fig. S2). Sequencing this amplicon showed that it corresponded to a gene encoding Ribulose-1,5-bisphosphate carboxylase/oxygenase (*Rca*). Further analysis confirmed that this sequence was missing from FN2122/2 (Notes S1). The gene *Rca1* was mapped to the middle of pea linkage group III on the genetic map of the JI281xJI399 RIL population, consistent with a location close to the *St* locus (Pellew & Sverdrup 1923, de Winton 1928). The *Rca1* gene was sequenced from JI2822 and JI1201 (an *st* mutant), but no lesion was found in the JI1201 sequence (Notes S1), suggesting that *Rca1* is unlikely to correspond to *Stipules reduced*.

No other AFLP fragment differences were observed, so we presumed that some fragments were masked by similar sized amplicons. We reasoned that any hidden polymorphisms would be revealed by sequencing RAD tags associated with PstI/MseI genomic DNA fragments. RAD libraries from the four genotypes were produced and 42bp sequence reads adjacent to the PstI site were generated (including the 3' terminal G of the PstI site). Sequences that did not begin with a G were ignored. This generated a total of ca. 25, 21.8, 22.9 and 42.8 million reads from FN1889-BC3, FN2002-BC3, FN2122-BC4 and JI2822 respectively. These sequences were collapsed, using the programme FASTX Collapser, into 480,671 individual unique RAD sequence tags shared with JI2822; the number of reads was recorded in the tag name. On average there were approximately 90 reads per tag, but there was a considerable range of read depths (Fig. S3). 331,668 tags occurred once, 55,481 had a read depth <10 and >1 and a smaller number of sequences were very common; 27 had a read depth >50,000 and were derived from chloroplast DNA or repetitive elements of the nuclear genome. Although most tags had a low read depth, most reads (ca. 40 million) were for tags with a read depth greater than 100, and fewer reads (ca. 2 million) were for tags with a lower read depth. Rare sequences were considered to be sequencing errors (Domoney *et al*., 2013) and were not investigated further.

Identical sequence tags in each of the four genotypes were identified with a simple Python script (Methods S1) and collated in an Excel sheet that recorded the sequence, its numerical identifier and read depth. For a given read depth from one genotype the read depth in the other three genotypes was over-dispersed with respect to a Poisson distribution: the variance was between 110 and 11,365 times greater than the mean. This suggested that, for

tags with low read depths, the absence of a tag from one genotype could be due to sampling variation. We therefore needed to identify a read depth where the signal for a missing sequence could be found among the noise of those absent by chance alone, as discussed by Domoney *et al*., (2013).

In order to screen for tags unlikely to be missing from FN2122/2 by chance alone a cut-off read depth of the tags needed to be determined. The frequency distribution of read depths (Fig. S3a) reflects the genomic copy number of the corresponding sequences; those present once per genome should be a common class. In a log-log plot of read depth vs the number of tags with that read depth, there is a shoulder in the plot between a read depth of ca. 100 and ca. 1000. A sample of sequences known to correspond to single copy genes that carry at least one PstI site was found to have tag read depths of 433 ± 272 (mean ± stdev, n = 34; Fig. S3b), consistent with these being in the shoulder between the read depths of 100 and 1000. There were 31,879 of these tags and a total of 9,913,708 reads in this group, giving an average read depth of 310 for single copy sequences. Analysis of JI2822 tags showed that for read depths less than 150, a high proportion were unique to that genotype (Domoney *et al.,* 2013). On average, sequence tags with a read depth of 100 for the *st* mutant (FN2122-BC4) corresponded to tags with a read depth of approximately 200 in JI2822, so a read depth of 150 was chosen as a compromise cut-off value between an excessive false discovery rate and a failure to detect genuinely missing sequences.

A total of 43,342 sequence tags were examined. These had a read depth of 150 or more in at least one of the samples representing the wild type *St* allele. Of these, 40,722 had a read depth of at least 150 in JI2822 and 40,020 were present in all samples (Fig. S4). These tags had a read depth of 92 ± 41 (mean ± stdev) in JI2822. 1,625 of these sequence tags were identified as being absent from the *st* mutant FN2122/2 and having at least 150 reads in at least one of the other three genotypes. Of these, 171 were missing from all the FN genotypes and 460 were absent from either FN1889/3 or FN2002/7. These 631 tags were eliminated from further analysis because both FN1889/3 and FN2002/7 carry a wild type *St* allele (Fig. S4). The remaining 994 tags had average read depths of 209, 95 and 111 reads in JI2822, FN1889/3 and FN2002/7 respectively. Although the number of tags identified in this way suggests a high false discovery rate, the probability that tags from both sides of a single PstI site are missing by chance alone is lower: the square of the false discovery rate. The identification of paired tags was therefore of interest.

The 994 tags absent from FN2122/2 *st* mutant reads, but present in all other genotypes, were used in a BLASTn search of transcript sequences downloaded from the

USDA pea unigene database (now available at https://www.coolseasonfoodlegume.org/sativum_unigene_v2). A set of 498 unigene contigs were identified that corresponded to these tags. This set was then used as the subject of a BLASTn query with all (including read depths <100) FN2122/2 sequence tags. Unigenes that matched one or more tags were eliminated from further investigation, because at least part of these genes is present in the *st* mutant FN2122/2. This eliminated 455 unigenes, leaving 43 for which no tag was found. These 43 unigenes were therefore potential candidates for *St* and were aligned against the medicago (*Medicago truncatula*) genome sequence (v3.5.1; Fig. S5).

Among these unigenes was *Rca* (contig19432), and a C2H2 zinc finger transcription factor (contig27619). In the medicago genome, the corresponding sequences were closely linked. The medicago C2H2 gene (Medtr5g080660) was annotated as *RABBIT EARS* (*RBE*)-*like* (Takeda *et al*., 2004; Krizek *et al*., 2006). Since the pea and medicago genomes align well (Duarte *et al*., 2014), the position of *St* on pea linkage group III was expected to correspond to the middle of medicago pseudomolecule 3. At first, this *Rca*-C2H2 region of the medicago genome appeared to correspond to a break in collinearity with the pea genome, however, a subsequent release of the medicago genome sequence (version 4.0 https://legumeinfo.org/genomes/gbrowse/Mt4.0) contains the sequence Medtr3g068095, which is a better match to contig27619. An *Rca* gene (Medtr3g068030) is located ca. 50kb distant from Medtr3g068095 and both lie in the expected syntenic region.

As described above, a presence/absence polymorphism for contig19432 (*Rca1*) existed, but there was no sequence polymorphism associated with *Rca* that distinguished *St* and *st* genotypes (Notes S1). A presence/absence polymorphism in wild type versus mutant was detected by PCR primers for contig27619, suggesting it was a better candidate for the *St* gene.

Confirmation of a candidate gene for *Stipules reduced*

Genomic sequences corresponding to contig27619 from JI2822 (*St*), JI813 (*St*) and JI143 (*st*), identified a SNP corresponding to an in-frame stop codon in the JI143 coding sequence. This was the only observed sequence difference between JI143 (with reduced stipules) and the other two genotypes (with normal stipules). Allele-specific PCR primers were designed to detect this SNP in a JI813xJI143 F2 population segregating for *st*. No recombination between the phenotype and SNP genotype was observed (Fig. S6).

Genomic sequences corresponding to contig27619 were obtained from a set of wild type and *st* mutant *Pisum* lines. All of the sequences from the mutant lines carried lesions in the predicted gene sequence as illustrated in Fig. 2 (see also Fig. S1 and Notes S2). Four

alleles resulted in a premature stop codon, two alleles had missense mutations at the 3' end of the coding sequence, and two alleles which failed to amplify any sequence from the gene were consistent with deletions.

The structure of the *St* Gene

The *St* gene is predicted to have a short upstream open reading frame (uORF) that is in frame with the main ORF. Of the two start ATGs, the main ORF ATG has a better fit to the Kozak consensus sequence for plants (Lütcke *et al*., 1987). There is a short intron predicted with respect to the unigene sequence (Notes S2), which was confirmed by PCR (Fig. S7). The predicted amino acid sequence (Notes S2) carries two recognisable domains. Towards the N terminus, beginning at C62, the sequence $CxxCx_{12}HxxxH$ corresponds to a C1–1iG family C2H2 zinc finger domain (Englbrecht *et al.*, 2004), while at the N terminus the sequence LDLELRL (beginning at L233) resembles an ERF-associated amphiphilic repression motif (EAR domain, Ohta *et al*., 2001).

Pisum is embedded in the genus *Lathyrus* (Schaefer *et al*., 2012) so three *Lathyrus* sequences were investigated; two were *L. odoratus* varieties (Lucy Hawthorne and Dorothy Eckford, with small stipules) and the third was *L. aphaca* (which has large stipules). The *L. odoratus* and *L. aphaca* nucleotide sequences were 95% and 94% identical to *St* and all three *Lathyrus* sequences contained an intron with a 4 bp insertion with respect to *Pisum*, furthermore *L. aphaca* carried a 3bp deletion in the coding sequence (Notes S2). The amino acid sequence of St is 93% identical to all three *Lathyrus* sequences, the two *L. odoratus* sequences are 99% identical to each other and 91% identical to *L. aphaca*. There were 5 residues that distinguished the large-stipuled species, pea and *L. aphaca*, from the small-stipuled species, *L. odoratus* and *M. truncatula* (Notes S2).

A BLASTp search using the predicted St amino acid sequence identified homologues in several sequenced legume species (Notes S2), and a closely related pea sequence (PsCam039889) was identified in the pea gene atlas (Alves-Carvalho *et al*., 2015, http://bios.dijon.inra.fr/FATAL/cgi/pscam.cgi). These sequences, together with the *Arabidopsis thaliana* RABBIT EARS were aligned and compared using MUSCLE and PHYLIP, with the purpose of determining the most likely orthologue of *Stipules reduced* in medicago, The zinc finger domains could be aligned unambiguously, permitting the construction of a phylogenetic tree (Fig. 3) which supports *St* as the orthologue of Medtr3g068095, and Medtr5g080660 as the orthologue of PsCam039889. A comparison between *St* and Medtr3g068095, the most closely related medicago nucleotide sequence, is shown in Fig. 4. There are regions of sequence conservation, notably including the zinc

finger and EAR domain coding sequences, but there is also an abundance of non-synonymous changes, in addition to the difference in intron structure. The medicago intron is 72 bp, compared to the 83 bp pea intron. The K_a/K_s ratio for the whole alignment is 1.29 (36/28) which is not significantly different from 1 ($\chi^2 = 1$, p=0.32). For the sequence following the zinc finger domain this value rises, ($K_a/K_s = 28/14 = 2$, $\chi^2 = 4.67$, p= 0.035) and at some locations Ka/Ks is in considerable excess of the mean for the whole sequence, indicative of diversifying selection (Fig. 4).

Considering the alignment as a whole (Fig. 4), there is no significant excess of amino acid differences in either of these two sequences, however the contribution from extra amino acids predicted in the Medtr3g068095 sequence is ignored. Both sequence conservation and diversification are therefore manifest in *St*.

Phenotypic description

The *st* mutation conditions a reduction in the size of the stipule (Fig. 1) and there are two mutant phenotypic classes (Pellew and Sverdrup, 1923; Apisitwanich and Swiecicki, 1992). With respect to wild type, the *st* stipule lamina is reduced by about 90% and *butterfly stipules* (st^{bs}) lamina by about 70% (Fig. S8). The least severe phenotypes are conditioned by mutations that affect the EAR domain in JI2653 and JI3521 (D234N) and JI3530 (L235P), with the D234N missense mutation in the st^{bs} alleles being the least severely affected.

Meicenheimer *et al*., (1983) concluded that the *st* mutant stipule phenotype derives from an early cessation in marginal meristematic activity. We were interested in whether this early cessation of cell division fully accounts for the reduction in organ size or whether there were also cell size differences. Accordingly, we measured cell size on the abaxial epidermis in four positions. The results (Table 1) show that cells are smaller in *st* mutants in all zones except for the proximal lateral sector, corresponding to the basal frill (mantle) in wild type, but only differences in the medial position were statistically significant. The difference in cell size between the medial and lateral positions in the wild type is significantly different (Table 1, Fig. S9), while there is no significant difference between these values for the *st* mutant. Therefore, reduction of cell size, particularly in the proximal medial region of the stipule, also contributes to the *st* phenotype.

A comparison of the vascularization of the *st* type allele (Fig. 5) shows that, while the vascular strands in the stem and the leaf of the *st* mutant have the configuration originally described for *Pisum* (Kupicha, 1975), the pattern of vascularization of the *st* stipule is less complex. For the *st* mutant, the main disturbance to vascularisation is within the basal

elaboration or mantle of the stipule (strand 7, Fig. 5). Apart from its effect on overall stipule size, the action of *St* is thus most apparent in the proximal part of the stipule.

Expression of *St*

In order to understand the process of stipule development further, we decided to ascertain whether the identity of the stipule affected *St* expression. The *coch* mutant, which replaces the stipule with a leaf-like structure, thus changing the identity of the organ at that position, allowed us to address this question. We analysed *St* expression of in a *coch* mutant and *Coch* expression in an *st* mutant by q-PCR (Fig. 6). This analysis shows that *St* transcript abundance is dependent on *Coch*, whereas *Coch* expression is unaffected by *St*, consistent with *St* expression being dependent on organ identity.

We further investigated the expression of *St* by *in situ* hybridisation (Fig. 7 and Fig. S10) in wild type and isogenic *coch* and *st* mutants. These results show that *St* is expressed in stipule primordia and developing stipules of wild-type plants. *St* expression was detected at plastochron 1 until plastochron 8 and was strongest on the adaxial side of the stipule (Fig. 7a), but weak in stipule vascular tissue (Fig. 7b). In young primordia the transcript appears in two symmetrically placed regions, presumably either side of the developing vasculature (Fig. S10c). The *St* transcript was absent from floral tissues (Fig. 7a, 7b), however a signal was detected on the flank of the inflorescence (Fig. 7a, b and d), consistent with expression in bracts.

No signal was detected in the *st* deletion mutant (Fig. 7e, 7f), which shows that the hybridization probe did not identify another *St*-related transcript. The *coch* mutant (Fig 7c, 7d) showed very weak *St* expression in stipules and bracts. This weak signal is consistent with the results of the q-PCR (Fig. 6).

Discussion

Methodology

Genome wide sequencing can detect mutations in mutant populations and so identify candidate genes in forward genetic screens (Tsai *et al*., 2011), but this depends on the availability of a reference genome sequence (Hwang *et al*., 2015, Campbell *et al*., 2016) which is not available for pea. Insertion mutagenesis can also tag genes facilitating their isolation (Schauser *et al.*, 1999, Tadege *et al*., 2008, Urbański *et al*., 2012), but in pea insertion mutagenesis is not available. Here we investigated an alternative approach in pea and demonstrated that RAD sequencing can identify sequences deleted from FN mutants.

The nature of mutations induced by ionising radiation depends on several factors, including the type and energy of the radiation and the cellular response to the free radical induced damage. When considering FN mutagenesis as a methodology for gene identification these factors need to be taken into account. The studies of Belfield *et al*. (2012) and Li *et al*. (2016) describe sequence variation associated with FN mutagenesis in Arabidopsis and rice respectively. Both studies attribute many types of mutation to FN mutagenesis, of which 36% were deletion mutations and 50 - 60% were single base substitutions. In Arabidopsis the deletions were small with only one greater than 55bp, while in rice 10% of the deletions were greater than 1kb and two (out of 873) were greater than 1Mb. These results contrast with our observations in pea, where no FN-induced allele (of 28 alleles distributed over 10 loci, Domoney *et al.* 2013, McAdam *et al*. 2017), was a single base change, suggesting that single base changes were relatively rarer in pea than in rice or Arabidopsis. It is notable that in these three examples the proportion and size of deletions increases with increasing genome size. The number of ways in which a deletion of x bp can occur, such that it that disrupts fewer than y genes, is a combinatorial function of intergenic distance, so it is perhaps not surprising that in pea, with a large genome and low gene density, that large deletions are more common.

The successful detection of the presence/absence of *St* depended on the large size of the FN-generated deletion. Previous studies in this population had shown that large deletions were common (Sainsbury *et al*., 2006, Wang *et al*., 2008, Hofer *et al*., 2009, Hellens *et al*., 2010, Moreau *et al*., 2012, Chen *et al*., 2012, Couzigou *et al*., 2012, Domoney *et al.*, 2013). The absence of at least two adjacent genes in FN2122/2 suggests a single large deletion has occurred in this line in the region encompassing both the *Rca* and *St* loci. Precedent for such a scale of deletion induced by FN in pea is the joint deletion of *alae keel- like* (*k*) and *Convicilin* (*Cvc*) (Domoney *et al*., 2013).

Our approach depended on reliable detection of a tag corresponding to a PstI site when it was actually present, so that any tag missing in a mutant would warrant further investigation. The variation in read depth of single copy sequences was very high and we found that a cut-off value of 150 reads was an adequate compromise between sensitivity and reliability. In JI2822, the *St* tags had read depths of 432 and 323 (Notes S2), while the *Rca* tags were 327 and 693 (Notes S1) consistent with the expected read depth of single copy genes (Fig. S3).

RNA-seq is an alternative approach (McAdam *et al*., 2017) which may be advantageous for large genome species such as pea, where many of the RAD-seq reads are effectively wasted because they derive from repetitive sequences not represented in the

transcriptome. However, genes involved in developmental patterning may be expressed in very few cells and therefore would be unusually rare in the transcriptome, so for these types of gene, the advantage of RNA-seq may fail to materialise. Furthermore, low abundance sequences would be most susceptible to stochastic loss.

The current lack of a genome sequence for JI2822 (the mutagenized line) hindered the identification of paired RAD tags flanking the same PstI site. For this reason, *M. truncatula*, the closest relative to pea for which genome sequence is available, was used in conjunction with pea transcriptome sequence data. The sequences of many of the RAD tags presumed missing from FN2122/2 corresponded to sequences distributed throughout the medicago genome (Fig. S5), as would be expected of tags missing by chance.

It has been estimated that each M2 from this FN population has, on average, 7 independent deletions (Domoney *et al*., 2013), suggesting that the BC4S1 individual studied here would not carry more than one deletion.

Publicly available transcriptome data were available for pea (Fransen *et al*., 2012, Kaur *et al*., 2012 and the USDA database at https://www.coolseasonfoodlegume.org/sativum_unigene_v2); but the work of Alves-Carvalho *et al*., (2015) was not available at the time this analysis was initially performed. Alignment of the RAD tags to the transcriptome sequences from the USDA database enabled the identification of paired sequences corresponding to the two sides of a PstI site, providing independent evidence for a deletion of the PstI site. This permitted the identification of a C2H2 zinc finger sequence as a candidate for the *St* gene, which was confirmed by the sequence analysis of independently obtained mutant alleles.

Analysis of soybean FN mutant populations has highlighted the advantages of other genome wide approaches such as resequencing or array hybridization when a reference genome sequence is available (Hwang *et al*., 2015, Campbell *et al*., 2016); these two studies also emphasise that in some cases simple deletions may not be the most frequent type of mutation. In both these cases genomic rearrangements rather than deletions were detected. It may be that the larger genome of pea, with interspersed repetitive elements, permits large deletions that are non-lethal. The approach we took, in this and previous studies (Chen *et al*., 2012, Hofer *et al*., 2009), screened for loss of PstI sites, so we may have missed rearrangements. Although our results demonstrate that a complete genome sequence of the target species is not required for this method of gene identification, RAD based deletion screens would be easier if extensive sequence were available.

The role of *Stipules reduced* in the pea compound leaf.

St is required for stipule enlargement (Meicenheimer *et al.*, 1983, Sinjushin *et al.*, 2011) rather than stipule identity, consistent with *St* being required for the elaboration of the basal frilled mantle. The reduced vascularisation of the *st* stipule may be a consequence of reduced stipule elaboration (Fig. 5) reminiscent of the reduced petal phenotype of the *rbe* mutant in *A. thaliana* (Takeda *et al*., 2004), *RBE* being the most closely related Arabidopsis sequence to *St* (over the whole length of the predicted amino acid sequence). In the *st* mutant there is no difference in cell size in the medial vs lateral position of the stipule, whereas there is a significant difference in wild type (Table 1, Fig. S9) showing that differences in cell expansion between zones of the stipule is dependent on *St*. Within the stipule, these *St*-dependent cell sizes may reflect medial vs lateral identity.

The st^{bs} mutant phenotype is weaker than the other *st* mutants (Fig. 1, Fig. S8) and results from a mis-sense (D234N) mutation within the C terminal EAR domain rather than non-sense mutation. The L235P substitution in JI3530 also occurs in the EAR domain, but the phenotype of this mutant is more severe than st^{bs} (Fig. 1) suggesting that this transcriptional repressor domain (Ohta *et al*., 2001) is required for *St* function. Like *rbe* (Huang *et al*., 2012, Huang & Irish, 2015, Li *et al*., 2016), all of the *st* alleles examined have reduced lamina growth at the proximal position of an organ.

St transcripts appear to be confined to stipules and bracts in pea but are not found in flowers, consistent with the lack of alterations to floral morphology in the *st* mutant. There has been no previous comment in the literature on an altered bract morphology in *st* mutants; this would be hard to detect because bracts are variable in size and frequency of appearance in pea. However, it has been noted previously that bracts are altered in *coch* mutants (Couzigou *et al*., 2012), so *Coch* is likely to be expressed in bract primordia, where it could up-regulate *St* expression and so determine the final size of bracts. A high level of *St* expression in the developing stipule (and bract) appears to be dependent on *Coch* (Figs 6 &7).

The *st* mutation in combination with *uni*, but neither mutant alone, completely abolishes stipule formation at upper nodes (Hofer *et al*., 2001, Kumar *et al*., 2009; 2013). The precise evolutionary relationship between the Arabidopsis C1–1iG family C2H2 zinc finger domain proteins and St is not clear, due to sequence duplications in Arabidopsis and possible recent diversifying selection acting on *St* (Fig. 4). RBE, through its regulation of TCP5 and microRNA164 (Huang & Irish, 2015) appears to be involved in regulating the switch between cell division and differentiation. *Uni* in pea leaves is responsible for a 'transient

phase of indeterminacy' (Hofer *et al*., 1997) which is manifest as continued meristematic activity in the leaf primordium while the *st* mutant has reduced stipule marginal meristem activity (Meicenheimer *et al*., 1983) so *St* promotes this marginal meristem activity. The complete loss of stipules, late in shoot development of the *st uni* double mutant, may reflect the roles of *Uni* in promoting primordial growth and *St* in promoting marginal growth.

The more distantly related Arabidopsis protein JAGGED (C1-1iA group, Englbrecht *et al*., 2004), like *St*, regulates cell growth and division (Dinney *et al*., 2004) and is involved in both bract and petal development. *St* regulates cell division to a greater extent than cell size, similar to JAGGED (Dinney *et al*., 2004) and to RABBIT EARS (Huang & Irish, 2015) more generally.

The very low level of *St* transcript in the *coch* mutant predicts that the *coch st* double mutant would be indistinguishable from *coch*. Yaxley *et al.,* (2001) reported that *coch st* and *coch* were indistinguishable, in disagreement with Blixt (1967), Marx (1987), Gourlay *et al.,* (2000) and Kumar *et al.,* (2009; 2013). Our transcript abundance results seem to be consistent with Yaxley *et al*., (2001), unless the small amount of *St* expression in the *coch* mutant can, under some circumstances, have consequences different from the null *st* mutant. The uORF may be relevant to these observations if it mediates posttranscriptional regulation (Laing *et al*., 2015), so the lower amount of *St* transcript in the *coch* mutant vs *Coch* (Fig. 6) may not necessarily result in a difference in the amount of St protein. Such regulation may be dependent on additional genetic or environmental factors and therefore explain the differences in the reported phenotypes of the *coch st* double mutant.

The *in situ* hybridisation and q-PCR results are in agreement. The *in situ* analysis additionally shows that the *St* transcript is limited to stipules and bracts, and is absent from floral meristems and other parts of the leaf primordium. The weak expression of *St* in the *coch* mutant is not associated with mis-location or mis-timing. It therefore appears that *Coch* is epistatic to *St*, consistent with *Coch* determining stipule identity. However, we cannot therefore completely rule out a role for *St* in determining stipule identity in certain genetic backgrounds because a leaf-like stipule structure was reported in an *af tl st* triple mutant (Gourlay *et al*., 2000), but this phenotype was noted to occur sporadically and only in the triple mutant. There is no evidence from the q-PCR result of a feedback between *St* and *Coch* whereby *St* would maintain *Coch* expression and indirectly stipule identity.

***Stipules reduced* in legume species.**

Coding sequence and structural differences between *St* and corresponding sequences in medicago and other legumes raises the possibility that *St* may have diverged in *Pisum*, in

association with the occurrence of large stipules. *L. aphaca* and *L. odoratus* differ in stipule size, the *L. aphaca* stipules being notably large, however sequence alignment (Notes S2) does not support a closer relationship between *Pisum* and *L. aphaca St* genes than *Pisum* and *L. odoratus St* genes. Nevertheless, one position (A199 see Notes S2) distinguishes the *L. aphaca* and *P. sativum* sequences from all the other sequences that were aligned in Fig. 3 and this could be targeted in future functional studies. Whether there is any association between stipule size and variation in the *St* gene more broadly in these taxa remains to be determined.

If *St* has undergone neo-functionalisation this may explain the elaboration of the pea stipule. There are strong signals of purifying selection acting on parts of the gene, yet in comparison to Medtr3g068095, some regions of *St* (and the gene as a whole) have an excess of amino acid substitutions given the nucleotide divergence (Fig. 4), suggestive of diversification (in one or other or both sequences). It should be noted that the vascularization of the stipules of these two *Lathyrus* taxa is different from each other and from pea (Kupicha, 1975), so these may represent three different consequences of *St* gene variants, or, stipule development in *Lathyrus* taxa may be independent of *St*.

Acknowledgements:

We thank Claire Domoney for her help and support, Clare Coyne for her advice on transcriptome sequence data, Wayne Powell for enabling these experiments, BBSRC the Welsh Government and the University of Aberystwyth for financial support.

Author contribution

CM, JH, ME, AS, MA, KS, TB, MH, VB, CF and NE planned and undertook experimental work and analysed data. MS wrote the Python script, advised on bioinformatics and edited the manuscript. NE, JH and AS wrote and edited the manuscript. All authors have read and accepted the manuscript.

References

Alves-Carvalho S, Aubert G, Carrère S, Cruaud C, Brochot A-L, Jacquin F, Klein A, Martin C, Boucherot K, Kreplak J, *et al*., 2015. Full-length de novo assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *The Plant Journal* **84**: 1–19.

Apisitwanich S, Swiecicki WK. 1992. A new allele, st^{bs}, at the *St* locus. *Pisum Genetics* **2**4: 13.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* 3: e3376.

Balanzà V, Martínez-Fernández I, Sato S, Yanofsky MF, Kaufmann K, Angenent GC, Bemer M, Ferrándiz C. 2018. Genetic control of meristem arrest and life span in Arabidopsis by a *FRUITFULL-APETALA2* pathway. *Nature Communications* **9**: 565.

Belfield EJ, Gan X, Mithani A, Brown C, Jiang C, Franklin K, Alvey E, Wibowo A, Jung M, Bailey K, Kalwani S, Ragoussis J, Mott R, and Harberd NP. 2012. Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*. *Genome Research* **22**: 1306–1315

Blixt S. 1967. Linkage studies in Pisum VII. The manifestation of the genes *cri*, and *coch*, and the double-recessive in *Pisum*. *Agric. Hort. Genet*. **25**: 131–144.

Campbell BW, Hofstad AN, Sreekanta S, Fu F, Kono TJY, O'Rourke JA, Vance CP, Muehlbauer GJ. Stupar RM. 2016. Fast neutron-induced structural rearrangements at a soybean *NAP1* locus result in *gnarled* trichomes. *Theor. Appl. Genet*. **129**:1725–1738.

Chen J, Moreau C, Liu Y, Kawaguchi M, Hofer J, Ellis N, Chen R. 2012. Conserved genetic determinant of motor organ identity in *Medicago truncatula* and related legumes. *Proc. Natl. Acad. Sci. USA* **109**: 11723 - 11728.

Couzigou J-M, Zhukov V, Mondy S, el Heba GA, Cosson V, Ellis THN, Ambrose M, Wen J, Tadege M, Tikhonovich I, et al. 2012. *NODULE ROOT* and *COCHLEATA* Maintain Nodule Development and Are Legume Orthologs of Arabidopsis *BLADE-ON-PETIOLE* Genes. *Plant Cell* **24**: 4498-4510.

Dash S, Campbell JD, Cannon EK, Cleary AM, Huang W, Kalberer SR, Karingula V, Rice AG, Singh J, Umale PE, Weeks NT, Wilkey AP, Farmer AD, Cannon SB. 2016. Legume information system (LegumeInfo. org): a key component of a set of federated data resources for the legume family. *Nucl. Acids Res*. **44**: D1181-D1188.

Domoney C, Knox M, Moreau C, Ambrose M, Palmer S, Smith P, Christodoulou V, Isaac PG, Hegarty M, Blackmore T et al. 2013. Exploiting a fast neutron mutant genetic resource in *Pisum sativum* (pea) for functional genomics. *Functional Plant Biology* **40**: 1261-1270.

Duarte J, Rivière N, Baranger A, Aubert G, Burstin J., Cornet L, Lavaud C, Lejeune-Hénaut I, Martinant J-P, Pichon J-P, et al. 2014. Transcriptome sequencing for high throughput SNP development and genetic mapping in Pea. *BMC Genomics* **15**:126.

Ellis THN, Davies DR, Castleton JA, Bedford ID. 1984 The Organization and Genetics of rDNA Length Variants in Peas. *Chromosoma* **91**: 74-81.

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ. 1998. Polymorphism of insertion sites of *Ty1-co*pia class retrotransposons and its use for linkage and diversity analysis in pea. *Mol. Gen. Genet* **260**: 9-19.

Englbrecht CC, Schoof H, Böhm S. 2004. Conservation, diversification and expansion of C2H2 zinc finger proteins in the *Arabidopsis thaliana* genome. *BMC Genomics* **5**: 39.

Franssen SU, Shrestha RP, Bräutigam A, Bornberg-Bauer E, Weber APM. 2011. Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. *BMC Genomics* **12**: 227.

Gourlay CW, Hofer JMI, Ellis THN. 2000. Pea Compound Leaf Architecture Is Regulated by Interactions among the Genes *UNIFOLIATA, COCHLEATA, AFILA,* and *TENDRIL-LESS*. *Plant Cell* **12**: 1279-1294.

Hellens R, Moreau C, Lin-Wang K, Schwinn KE, Thomson SJ, Fiers MWEJ, Frew TJ, Murray SR, Hofer JMI, Jacobs JME, et al. 2010. Identification of Mendel's White Flower Character *PLoS ONE* **10**: e13230.

Hofer JMI, Gourlay CW, Ellis THN. 2001. Genetic control of leaf morphology: A partial view. *Annals of Botany* **88**: 1129-1139.

Hofer J, Turner L, Hellens R, Ambrose M, Matthews P, Michael A, Ellis N. 1997. *Unifoliata* regulates leaf and flower morphogenesis in pea. *Curr. Biol*. **7**: 581-587.

Hofer J, Turner L, Moreau C, Ambrose M, Isaac P, Butcher S, Weller J, Dupin A, Dalmais M, Le Signor C, et al. 2009 *Tendril-less* regulates tendril formation in pea leaves. *Plant Cell* **21**:420-428.

Huang T, Irish VF. 2015. Temporal Control of Plant Organ Growth by TCP Transcription Factors. *Curr. Biol.* **25**: 1765-1770.

Huang T, López-Giráldez F, Townsend JP, Irish VF. 2012. RBE controls microRNA164 expression to effect floral organogenesis. *Development* **139**: 2161-2169.

Hwang WJ, Kim MY, Kang YJ, Shim S, Stacey MG, Stacey G, Lee S-H. 2015. Genome-wide analysis of mutations in a dwarf soybean mutant induced by fast neutron bombardment *Euphytica* **203**: 399–408.

Kaur S, Pembleton LW, Cogan NOI, Savin KW, Leonforte T, Paull J, Materne M, Forster JW. 2012. Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers *BMC Genomics* **13**:104.

Knox M, Moreau C, Lipscombe J, Baker D, Ellis N. 2009. High-throughput retrotransposon-based fluorescent markers: improved information content and allele discrimination. *Plant Methods* **5**: 10.

Krizek BA, Lewis MW, Fletcher JC. 2006. *RABBIT EARS* is a second-whorl repressor of *AGAMOUS* that maintains spatial boundaries in Arabidopsis flowers. *Plant J.* **45**: 369-383.

Kumar S, Mishra RK, Kumar A, Srivastava S, Chaudhary S. 2009. Regulation of stipule development by *COCHLEATA* and *STIPULE-REDUCED* genes in pea *Pisum sativum. Planta* **230**: 449–458.

Kumar A, Sharma V, Kumar S. 2013. Interaction between *cochleata* and *stipule-reduced* mutations results in exstipulate hypertrophied leaves in *Pisum sativum L. Ind. J. Exptl. Biol.* **51**: 492-501.

Kupicha FK. 1975. Observations on the vascular anatomy of the tribe Vicieae (Leguminosae) *Bot. J. Linn. Soc.* **70**: 231-242.

Laing WA, Martínez-Sáncheza M, Wright MA, Bulley SM, Brewster D, Dare AP, Rassam M, Wang D, Storey R, Macknight RC et al. 2015. An Upstream Open Reading Frame Is Essential for Feedback Regulation of Ascorbate Biosynthesis in Arabidopsis. *Plant Cell*. **27**: 772-786.

Lamprecht H. 1960. Eine neue Stipel-Mutante bei *Pisum*. *Agr. Hort. Genet*. **18**: 209-213

Lewis GP, Schrire B., Mackinder B, Lock M. 2005. *Legumes of the World*. Kew Publishing.

Li G, Chern M, Jain R, Martin JA, Schackwitz WS, Jiang L, Vega-Sánchez ME, Lipzen AM, Barry KW, Schmutz J and Ronald PC. 2016. Genome-wide sequencing of 41 rice (*Oryza sativa L*.) mutated lines reveals diverse mutations induced by fast-neutron irradiation. *Molecular Plant* **9**: 1078–1081

Li J, Wang Y, Zhang Y, Wang W, Irish VF, Huang T. 2016. *RABBIT EARS* regulates the transcription of TCP4 during petal development in Arabidopsis. *J. Exp. Bot.* **67**: 6473-6480.

Lütcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA. 1987. Selection of AUG initiation codons differs in plants and animals. *EMBO J.* **6**: 43–48.

McAdam SAM, Eléouët MP, Best M, Brodribb TJ, Murphy MC, Cook SD, Dalmais M, Dimitriou T, Gélinas-Marion A, Gill WM, Hegarty M, Hofer JMI, Maconochie M, McAdam EL, McGuiness P, Nichols DS, Ross JJ, Sussmilch FC, Shelley Urquhart S. 2017. Linking Auxin with Photosynthetic Rate via Leaf Venation. *Plant Physiol.* **175**: 351-360.

Marx GA. 1987. A suite of mutants that modify pattern formation in pea leaves. *Plant Mol. Biol. Rep*. **5**: 311–335.

Meicenheimer RD, Muehlbauer FJ, Hindman JL, Gritton ET. 1983. Meristem characteristics of genetically modified pea (*Pisum sativum*) leaf primordia. *Can. J. Bot.* **61**: 3430-3437.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*. **17**: 240-248.

Moreau C, Ambrose MJ, Turner L, Hill L, Ellis THN, Hofer JMI. 2012. The *b* gene of pea encodes a defective flavonoid 3',5'-hydroxylase, and confers pink flower color. *Plant Physiol.* **159**: 759-68.

Ohta M, Matsui K, Hiratsu K, Shinshi H, Ohme-Takagi M. 2001. Repression Domains of Class II ERF Transcriptional Repressors Share an Essential Motif for Active Repression *Plant Cell* **13**: 1959–1968.

Pellew C, Sverdrup A. 1923. New observations on the genetics of peas. *J. Genetics* **13**: 125-131.

Perrier X, Flori A, Bonnot F. 2003. Data analysis methods. *in*: **Hamon, P., Seguin, M., Perrier X, Glaszmann JC.** *eds*., Genetic diversity of cultivated tropical plants. Enfield, Science Publishers. Montpellier. pp 43-76.

Sainsbury F, Tattersall AD, Ambrose MJ, Turner L, Ellis THN, Hofer JMI. 2006. A *crispa* null mutant facilitates identification of a *crispa*-like pseudogene in pea. *Functional Plant Biology* **33**: 757–763.

Schaefer H, Hechenleitner P, Santos-Guerra A, Sequeira MM de, Pennington RT, Kenicer G, Carine MA. 2012. Systematics, biogeography, and character evolution of the legume tribe Fabeae with special focus on the middle-Atlantic island lineages. *BMC Evol. Biol.* 12: 250.

Schauser L, Roussis A, Stiller J, Stougaard J. 1999. A plant regulator controlling development of symbiotic root nodules *Nature* **402**: 191-195.

Sidorova KK. 1975. Induced pea mutant with creeping stem. *Pisum Newsletter* **7**: 57–58

688 **Sinjushin AA, Khartina GA, Gostimskii SA. 2011.** New allele of the *COCHLEATA* gene in
689 pea *Pisum sativum* L. *Russ. J. Genet*. **47**: 1422-1427.
690 **Tadege M, Wen J, He J, Tu H, Kwak Y, Eschstruth A, Cayrel A, Endre G, Zhao PX,**
691 **Chabaud M, et al. 2008.** Large-scale insertional mutagenesis using the *Tnt1* retrotransposon
692 in the model legume *Medicago truncatula. Plant J.* **54**: 335–347.
693 **Takeda S, Matsumoto N, Okada K. 2004.** *RABBIT EARS*, encoding a SUPERMAN-like
694 zinc finger protein, regulates petal development in *Arabidopsis thaliana*. *Development* **131**:
695 425-434.
696 **Tsai H, Howell T, Nitcher R, Missirian V, Watson B, Ngo KJ, Lieberman M, Fass J,**
697 **Uauy C, Tran RK, et al. 2011.** Discovery of Rare Mutations in Populations: TILLING by
698 Sequencing *Plant Physiology* **156**: 1257–1268.
699 **Urbański DF, Małolepszy A, Stougaard J, Andersen SU. 2012.** Genome-wide *LORE1*
700 retrotransposon mutagenesis and high-throughput insertion detection in *Lotus japonicus.*
701 *Plant J.* **69**: 731–741.
702 **Vassileva M. 1979.** Induced mutagenesis in *Pisum*. I. Genetic studies on the *acacia* mutant.
703 *Genet. Sel.* **12**: 396-408.
704 **Wang Z, Luo Y, Li X, Wang L, Xu S, Yang J, Weng L, Sato S, Tabata S, Ambrose M,**
705 **Rameau C, et al. 2008.** Genetic control of floral zygomorphy in pea (*Pisum sativum* L.)
706 *Proc. Natl. Acad. Sci. U.S.A* **105**: 10414-10419.
707 **de Winton D. 1928** Further linkage work in *Pisum sativum* and *Peimula sinensis Z. Ind.*
708 *Abst. -u Vererb. -lehre, Suppl.* **2**: 1594-1600.
709 **Yaxley JL, Jablonski W, Reid JB. 2001.** Leaf and Flower Development in Pea (*Pisum*
710 *sativum L.*): Mutants *cochleata* and *unifoliata*. *Annals of Botany* **88**: 225-234.
711 **Zhuang L, Ambrose M, Rameau C, Weng L, Yang J, Hu X, Luo D, Li X. 2012.**
712 *LATHYROIDES*, encoding a *WUSCHEL*-related homeobox1 transcriptional factor, controls
713 organ lateral growth and regulates tendril and dorsal petal identities in garden pea (*Pisum*
714 *sativum* L.). *Molecular Plant*. **5**: 1333–1345.
715

Supplementary Information Legends:

Fig. S1

The stipule phenotypes of pea lines carrying *St* or *st* alleles is illustrated for corresponding mutant and wild type lines.

Fig. S2

A screen-shot of AFLP profiles using the selective primers Pst+AG, and Mse+GTC that identified a presence absence polymorphism of an amplicon approximately 212bp in size present among the products of the pea line JI2822, but absent from the isogenic FN2122/2 products is shown.

Fig. S3

The frequency distribution of read depths of RAD-tags in the pea line JI2822 is presented.

Fig. S4

A Venn diagram presents the number and pattern of the distribution of RAD-tags among the four pea lines examined.

Fig. S5

The positions of *M. truncatula* homologues of the 43 candidate pea genes on the 8 *M. truncatula* pseudomolecules (Mt Assembly v3.5.1) is presented.

Fig. S6

A co-segregation analysis of *St* and its candidate gene in an F2 population of the cross between the pea lines JI143 (*st*) and JI813 (*St*) as assessed with allele specific PCR primers is presented.

Fig. S7

A PCR analysis confirms the presence of an intron in the pea *St* gene.

Fig. S8

The stipule area in selected mutant and wild type pea plants is presented.

Fig. S9

Stipule epidermal cell area is plotted for F2 segregants of the cross between the pea lines Flagman (*St*) and Filby (*st*).

Fig. S10

The *St* transcript is localised by *in situ* hybridisation to thin sections and whole mount preparations of pea tissue from wild type and *coch* mutant lines, using the *st* deletion mutant as a control for cross hybridisation with related transcripts.

Methods S1

This file is the Python script used to collate, count and record identical short sequence tags.

Notes S1

This file contains information relevant to the sequence analysis and alleles of the pea *Rca* gene.

Notes S2

This file contains information relevant to the sequence analysis, alleles, and phylogenetic relationships of the pea *Stipules-reduced* gene.

Table 1 Epidermal cell areas in wild type and *st* mutant pea plants

	Proximal		Distal	
Genotype	*St*	*st*	*St*	*st*
medial	4.1 ± 0.9	3.0 ± 0.5**	4.1 ± 1.3	3.2 ± 0.8*
lateral	2.8 ± 0.6	3.1 ± 0.8	3.5 ± 0.6	3.0 ± 0.7
|medial - lateral|	1.3	0.1	0.6	0.2
t	9.15**	0.98	5.85**	1.45

The cell area of stipule epidermal pavement cells area was measured at four locations in *St* and *st* F2 segregants (n = 30 for each class) of the cross between the pea lines Flagman (*St*) and Filby (*st*). Epidermal cell areas (in units of 1,000 μm^2) are given as Mean ± standard deviation n = 30. Measurements for *st* that are significantly different from wild type (Kruskal-Wallis test) are marked with asterisks (*, $p < 0.05$; **, $p < 0.01$). For the difference between the means Student's t values are given. See Fig. S9 for a graphical representation of these data.

Figure Legends

Figure 1. The classes of *stipules reduced* mutant phenotypes

a: The pea wild type (JI2822) and corresponding FN induced deletion mutant (FN2122/2) are shown in two views; a single compound leaf, adaxial view on the left, a whole shoot tip in side view on the right. The single leaves comprise a proximal pair of stipules at the base of the leaf, a pair of leaflets and a distal tendril (obscured by the leaflets). These illustrate the small stipule size of a *st* null allele. The shoot tips illustrate the difference in the way the shoot apex is enclosed by stipules in *St* vs *st*.

b: Single compound leaves from the wild type progenitor (JI3538), left, and its corresponding weak mutant allele (JI2653), designated st^{bs} (*stipules reduced - butterfly stipules*).

Plants were 1 month old. The scale bar indicates 5 centimetres. Additional alleles are shown in Notes S2.

Figure 2. Gene structure

The structure of the pea *St* gene is illustrated. Open reading frames, including a predicted short upstream open reading frame are boxed. The zinc finger and EAR domains are shaded in black, and the intron is marked as a line. The nature and location of observed mutations is given above with the corresponding JI accession numbers for the lines carrying each mutation given below. The scale below is in nucleotides.

Figure 3 Phylogenetic analysis

Phylogenetic relationship between *Stipules reduced*-like amino acid sequences in selected legume species. The tree was generated from the sequence alignment of the zinc finger domain as shown in Notes S2. Bootstrap values in excess of 40% are given to the left of the corresponding branch. Groupings of sequences are largely congruent with species phylogeny as indicated by colour shading. Open arrows indicate the most closely related sequences deduced from the *M. truncatula* genome sequence and filled arrows indicate the pea *Stipules reduced* and its closest relative in *Pisum* (PsCam039889). Abbreviations are: Ardu *Arachis duranensis*, Araip *Arachis ipaensis*, At *Arabidopsis thaliana*, Ca_D, *Cicer arietinum* var. Desi, Ca_K *Cicer arietinum* var. Kabuli, Ccaj *Cajanus cajan*, Gm *Glycine max*, Lan *Lupinus angustifolius*, Lj *Lotus japonicus*, Mt *Medicago truncatula*, Pv *Phaseolus vulgaris*, Tp *Trifolium pratense*, Va *Vigna angularis*, Vr *Vigna radiata*

Figure 4 Comparison between *Stipules reduced* and Medtr3g068095

Analysis of pea *Stipules reduced* and *M.truncatula* Medtr3g068095 sequences.

a: The Ka/Ks ratio (y axis) is plotted for a sliding window of 25 residues of the alignment. Where Ks is zero the line is broken. Dashed line, Ka=Ks. Dotted line represents mean + 3x Std. Dev. of Ka/Ks.

b: The location of a gap caused by a difference in splice donor sites is marked and highlighted with diagonal stripes. The conserved zinc finger and EAR domains are marked, and, together with other conserved sequences, are highlighted with a grey background.

Note the x axis does not represent actual positions in either sequence because the alignment includes indels.

Figure 5 Stipule vascular supply

Phloroglucinol stained vascular strands of pea (**a**) wild type and (**b**) *stipules reduced* stipules. Diagrams to the left follow the vasculature numbering system of Kupicha (1975), these numbers are shown as labels on acid-macerated samples, right.

Figure 6 Expression of *Stipules reduced* and *Cochleata* in wild type and mutant plants

Relative expression levels of *St* and *Coch* in wild type, single and double mutant pea seedling shoot apices were measured by q-PCR. Relative expression levels and the experimental error estimated from three independent plants measured in triplicate is shown. The wild type *St* allele is from JI2822, the *coch* allele is the deletion mutation of FN3185 generated in the JI2822 background (Couzigou *et al*., 2012) and the *st* allele is the deletion allele of FN2122 in the JI2822 background, which serves as a negative control indicative of absence of the *St* transcript.

Figure 7 In situ hybridization of *Stipules reduced* to wild type and mutant sections.

RNA *in situ* hybridisations with an *St* probe. All stained pea tissues are stipules except for the three arrowed locations which are in the position of bracts. Apical meristems are adjacent to the symbol x. A) Wild type (JI2822) longitudinal section. B) Wild type (JI2822) transverse section. C) *coch* mutant (FN3185) longitudinal section. D) *coch* mutant (FN3185) transverse section, the asterisk marks a weakly stained stipule. E) *st* mutant (FN2122) longitudinal section. F) *st* mutant (FN2122) transverse section. The FN2122 *st* mutant acts as a negative control because the *St* gene, and therefore its transcript, is missing from plants of this genotype. A 200 μm scale bar is given in each panel.